

基于深度学习的多模态融合三维人脸识别^①



胡乃平, 贾浩杰

(青岛科技大学 信息科学技术学院, 青岛 266061)

通信作者: 贾浩杰, E-mail: jhjqust@163.com

摘要: 二维人脸识别受光照、遮挡和姿态的影响较大. 为了克服二维人脸识别的缺点, 本文提出了一种基于深度学习的多模态融合三维人脸识别算法. 该方法首先使用卷积自编码器将彩色图像和深度图进行融合, 将融合后的图像作为网络的输入进行预训练, 并且设计了一种新的损失函数 cluster loss, 结合 Softmax 损失, 预训练了一个精度非常高的模型. 之后使用迁移学习将预训练的模型进行微调, 得到了一个轻量级神经网络模型. 将原始数据集进行一系列处理, 使用处理之后的数据集作为测试集, 测试的识别准确率为 96.37%. 实验证明, 该方法弥补了二维人脸识别的一些缺点, 受光照和遮挡的影响非常小, 并且相对于使用高精度三维人脸图像的三维人脸识别, 本文提出的算法速度快, 并且鲁棒性高.

关键词: 三维人脸识别; 多模态融合; 深度学习; 卷积神经网络; 损失函数; 迁移学习

引用格式: 胡乃平, 贾浩杰. 基于深度学习的多模态融合三维人脸识别. 计算机系统应用, 2022, 31(8): 152-159. <http://www.c-s-a.org.cn/1003-3254/8687.html>

3D Face Recognition with Multi-modal Fusion Based on Deep Learning

HU Nai-Ping, JIA Hao-Jie

(School of Information Science and Technology, Qingdao University of Science and Technology, Qingdao 266061, China)

Abstract: Two-dimensional (2D) face recognition is greatly affected by illumination, occlusion, and attitude. To overcome these shortcomings, this study proposes a 3D face recognition algorithm with multi-modal fusion based on deep learning. Firstly, the convolutional autoencoder fuses the color image and the depth map, and the fused image is input to the network for pre-training. In addition, a new loss function cluster loss is designed for pre-training in combination with the Softmax loss, so as to obtain a highly accurate model. Then, transfer learning is employed to fine-tune the pre-trained model, and thus a lightweight neural network model is obtained. The processed original dataset is used as the test set, and the identification accuracy of the test reaches 96.37%. Experimental results verify that the proposed method makes up for some shortcomings of 2D face recognition, and it is less affected by illumination and occlusion. Compared with 3D face recognition using high-precision 3D face images, the proposed algorithm is faster and more robust.

Key words: 3D face recognition; multi-modal fusion; deep learning; convolutional neural network (CNN); loss function; transfer learning

近年来, 得益于数据集的日益庞大和硬件设备日新月异的发展, 深度学习应用越来越广泛, 带动了人脸识别的快速发展. 目前二维人脸识别已趋于饱和, 在光照统一、表情一致的实验室条件下, 识别准确率已经

非常高. 但是二维人脸识别受光照、遮挡和姿态等的影响较大, 在光线昏暗、面部遮挡的情况下, 不能正确地识别出人脸.

三维人脸除了包含彩色图的纹理信息之外, 也包

^① 收稿时间: 2021-10-16; 修改时间: 2021-12-14; 采用时间: 2022-01-28; csa 在线出版时间: 2022-06-16

含了深度信息. 相对于二维人脸识别, 三维人脸识别应对遮挡、光照和姿态的变化更具有优势. 因此, 三维人脸识别适合更多的场景. 三维人脸识别根据数据来源的不同可分为3类: 基于RGB图像的三维人脸识别、基于高质量三维扫描图像的三维人脸识别、基于低质量RGB-D图像的三维人脸识别. 基于RGB图像的三维人脸识别通过RGB彩色图像重建三维人脸模型, 但是重建模型较为复杂, 且参数极多速度较慢. 基于高质量图像的三维人脸识别使用高质量三维人脸数据进行识别, 但是获取高质量的三维人脸数据代价高昂, 需要特定的高精度扫描仪进行采集, 并且采集的数据非常大, 会导致模型过于复杂. 目前高精度三维人脸数据集有Bosphorus^[1]和BU-3DFE^[2]等, Bosphorus为点云数据, BU-3DFE为网格数据. 基于低质量RGB-D图像的三维人脸识别采用的数据来源于RGB-D相机, 这类相机能够同时采集彩色图和深度图, 速度快且成本较低. 但是低质量的深度图存在大量的空洞、毛刺, 造成识别困难. Mu等人^[3]单独使用低质量深度图训练神经网络, 得到的模型训练准确度较低. 这说明了单独使用低质量深度图进行人脸识别难以进行有效的区分. 因此本文使用基于低质量RGB-D图像的三维人脸识别方法, 结合彩色图和深度图进行人脸识别, 弥补单独使用低质量深度图的不足. 低质量人脸深度图的数据集有Lock3DFace^[4]、EURECOM^[5]和Texas3DFRDatabase^[6]等, 深度图数据集都含有对应的彩色图. 本文使用的数据集是Texas3DFRDatabase.

1 相关工作

本节介绍基于高质量图像的三维人脸识别方案, 基于低质量RGB-D图像的三维人脸识别方案和人脸识别损失函数.

基于高质量图像的三维人脸识别. 随着技术的进步, 学者逐步转向三维人脸识别, 但是高质量三维人脸数据库数量依然比较少. Gilani等人^[7]在现有高质量三维人脸数据的基础上, 通过不同图像之间的组合来合成新的人脸数据, 并且得到不同视角下的点云数据, 用来训练CNN. Cai等人^[8]在三维人脸数据上选取眼睛和鼻尖3个点进行姿态校准. 之后从三维人脸图像中提取4个子图像, 4个子图具有一定的重叠, 把子图分别输入到CNN中, 得到4个特征向量串联进行识别. Kim等人^[9]将单一的三维人脸数据合成多个不同表情

的人脸数据, 之后对二维人脸识别模型进行迁移学习, 少量的数据得到不错的效果.

基于低质量RGB-D图像的三维人脸识别. RGB-D图像包括彩色图和深度图. 单独使用深度图进行识别任务不能得到较好的效果, 因此将彩色图和深度图进行融合会得到更好的识别结果. 融合的方式分为3种: 信号层融合、特征层融合、决策层融合^[10]. 信号层融合即在原始图像上进行融合. Kusuma等人^[11]使用主成分分析(PCA)在信号层方面融合了二维人脸图像和三维人脸图像, 并探讨了不同模态数据之间的依赖关系. Jiang等人^[12]将二维数据的3通道和三维数据的3通道组合成6通道的数据输入到卷积神经网络中进行识别. 特征层融合即将不同模态数据的特征进行融合. Lee等人^[13]使用两路神经网络分别提取彩色图和深度图的特征, 并进行融合. 并且针对训练数据不够的问题进行了迁移学习. Li等人^[14]利用多通道稀疏编码进行不同模态之间的人脸对齐, 提取不同模态的人工特征进行特征融合. 决策层融合分别对不同模态的数据进行识别, 将对应得分通过策略进行融合. Chang等人^[15]使用主成分分析法对彩色图和深度图进行识别, 将两个得分进行加权融合. Cui等人^[16]对3种融合进行了对比研究, 并且提出了一种基于特征层和决策层的混合融合, 取得了当时最先进的性能.

人脸识别损失函数. 通常来说, 识别或分类任务使用Softmax损失函数, 但是Softmax损失函数只能保证目标具有可分性, 对人脸识别这种类间相似性较大的任务来说, 具有非常大的局限性. 因此, 损失函数是人脸识别方向的研究热点. 人脸识别损失函数主要有两种思路: 减少类内差异, 增大类间可分离性. 基于第1种思路, Wen等人^[17]提出了中心损失(center loss), 惩罚了样本与其中心的欧氏距离, 将相同的类聚集在类特征中心周围, 减少了类内差异, 结合Softmax损失函数, 取得了不错的识别效果. 基于第2种思路, Deng等人^[18]提出角边缘损失, 将角度约束加入到Softmax损失函数中, 极大地提高了类间可分离性, 能够获得人脸的高分辨率特征. Cai等人^[19]结合两种思路, 提出了岛屿损失(island loss), 顾名思义, 使相同的类聚集在一起像岛屿一样, 减少类内变化的同时扩大类间的差异, 性能得到显著提升.

本文的主要贡献如下:

(1) 设计自动编码器将彩色图和深度图在特征层

面进行融合. 得到融合图像作为识别任务的输入.

(2) 基于巴氏距离的思想, 提出一种新的损失函数 cluster loss. 结合 Softmax 损失进行识别任务, 不仅能够减少类内的变化, 而且能够扩大类间的差异.

(3) 使用迁移学习, 将融合图像训练得到的模型使用彩色图像进行微调. 得到的新模型使用彩色图像作为输入, 依旧可以克服光照和面部遮挡这一缺点. 将人脸彩色图像模拟戴口罩和变暗, 识别率依然较高.

2 算法

2.1 多模态融合

多模态融合指融合不同形式的数据. 本文需要融合的数据是人脸的彩色图和深度图. 使用卷积自动编码器对两种不同形式的图像进行融合. 卷积自动编码器是将传统的自编码器的无监督学习方式, 结合了卷积层、池化层、上采样层或者转置卷积层等操作. 图

像经过解码层得到特征向量, 特征向量经过编码层得到重构图像. 自编码器自动从样本中学习相关特征而不是通过人工进行提取.

卷积自编码器网络结构如图 1 所示. 网络分为编码层和解码层. 编码层包括 4 组卷积层+池化层和两组全连接层. 解码层包括 4 组卷积+上采样层和一个卷积层. 原始图像 r 和 d 分别是同一人脸图像的彩色图和深度图, 具有互补的关系. 彩色图和深度图同时作为卷积自编码器输入, 经过编码层之后, 得到两组 1 504 维的向量. 将两组 1 504 维的向量进行串行结合得到 3 008 维的向量, 并将融合后的向量和两组 1 504 维的向量作为解码层的输入. 经过解码层之后, 两组 1 504 维的向量会分别输出重构图像, 融合向量会输出融合图像. 公式表示如下:

$$\begin{cases} \min(r - D_r(E_r(r))) \\ \min(d - D_d(E_d(d))) \end{cases} \quad (1)$$

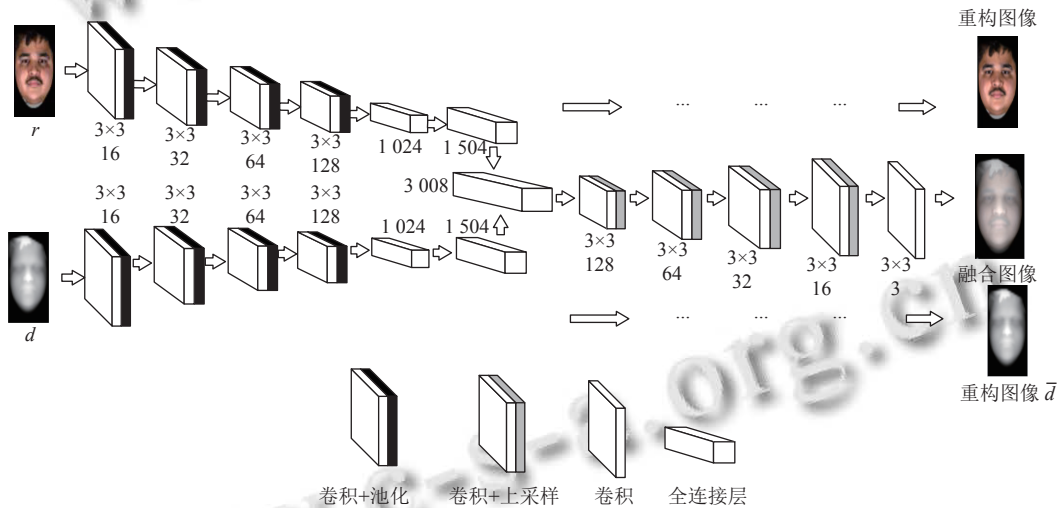


图 1 卷积自编码器网络结构

以图像 r 为例, E_r 是图像 r 的编码过程, 从 r 中提取特征, D_r 是图像 r 的解码过程, 通过特征解码出重构图像 \bar{r} . 图像 d 亦是如此. 图像 r 和 d 分别学习各自的特征, 在编码层末端将特征进行串联, 解码层共享权重, 实现不同模态特征的联合统一表达. 损失函数是原始图像与重构图像之间的差值, 损失函数公式表示如下:

$$loss = \min \left(\frac{1}{m} \left(\sum_{i=1}^m \frac{1}{2} (\|r^i - \bar{r}^i\|^2 + \|d^i - \bar{d}^i\|^2) \right) \right) \quad (2)$$

其中, m 表示共有 m 组数据, r^i 和 d^i 分别表示原始的彩

色图和深度图, \bar{r}^i 和 \bar{d}^i 表示重构的彩色图和深度图.

在使用卷积自编码器进行多模态数据融合时, 有以下几点说明:

(1) 不能将全连接层去掉或者使用卷积层代替, 否则编码器不能学习到有用的特征, 图像无法正常生成. 全连接层的作用是打乱空间结构. 全连接层可用 1×1 的卷积网络去代替, 1×1 的卷积网络也可打乱空间结构, 并且参数比全连接层少, 速度更快, 缺点是重构图像效果没有使用全连接层的效果好.

(2) 解码层的卷积+上采样层可用转置卷积

(transposed-convolution)^[20]代替, 转置卷积为卷积的反向操作, 可以使低分辨率图像向高分辨率转换. 但是文献 [21] 中指出使用 convolution+upsample 组合, 较之于 transposed-convolution, 可以更好地生成图像细节, 并且运算速度相当. 因此本节算法使用 convolution+upsample 组合.

2.2 损失函数

人脸识别或者表情识别存在着较高的类间相似性, 导致识别性能显著下降. 传统的卷积神经网络进行人脸识别任务时, 使用 Softmax loss 作为损失函数, 惩罚分类错误的样本, 能够分离出不同类别的特征, 但是特征向量是分散的, 并且会出现不同类别的特征重叠的情况, 如图 2(a) 所示. Wen 等人^[17] 提出了 center loss, 结合 Softmax loss, 实现了类内紧致性, 计算每一个类的中心, 并惩罚了每一类的样本和类中心之间的距离. 使特征向量聚集在一起, 但是 center loss 没有考虑到类间相似性, 依然会出现特征重叠的情况, 如图 2(b) 所示. 本节提出了一种新的损失函数 cluster loss, 能够提高人脸特征的判别能力. 惩罚不同类别的类中心之间的相似度, 将不同的类别远远推开. 结合 Softmax loss, 不仅实现了类内紧致而且能够扩大类间差异, 如图 2(c) 所示. 在本节首先简要回顾一下 center loss, 然后重点介绍我们提出的损失函数 cluster loss.

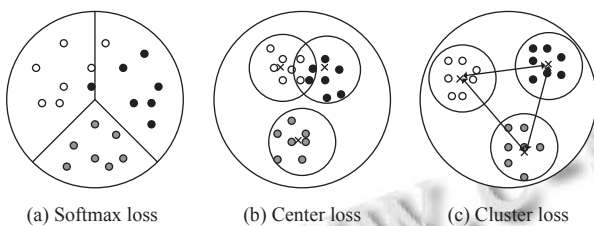


图2 3种损失函数示意图

2.2.1 Center loss

Center loss 为特征向量与类中心的距离的平方和, 公式表示为式 (3):

$$L_c = \frac{1}{2} \sum_{i=1}^m \|x_i - c_{y_i}\|^2 \quad (3)$$

其中, y_i 为样本 x_i 对应的类别, c_{y_i} 为类别 y_i 的类中心. 前向传播过程中, 损失函数为联合 Softmax loss 和 center loss 的加权和, 公式表示为式 (4). λ 为平衡因子, 用来平衡两种损失函数.

$$L_{s+c} = L_s + \lambda L_c = - \sum_{i=1}^m \log \frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^n e^{W_j^T x_i + b_j}} + \frac{\lambda}{2} \sum_{i=1}^m \|x_i - c_{y_i}\|^2 \quad (4)$$

2.2.2 Cluster loss

如图 2(b) 所示, 单纯使用 center loss 不能避免特征重叠的情况, 为了解决这个问题, 我们提出了一个新的损失函数 cluster loss, 能够扩大类间差异. 使用巴氏距离定义 cluster loss, 并用来衡量两个向量之间的距离. Cluster loss 公式表达为式 (5):

$$L_{\text{cluster}} = L_c + \sum_{j \in N} \left(-\ln \sum_{\substack{i \in N \\ i \neq j}} \sqrt{c_j c_i} \right) \quad (5)$$

其中, 第 1 项为 center loss, 用来惩罚样本与对应类中心之间的距离, 第 2 项表达式用来减少不同的类中心的相似性, 以扩大类间距离. N 为类别标签集合, i 和 j 分别代表不同的类别, c_i 和 c_j 分别是 i 和 j 对应的类中心. 前向传播过程中, 整体损失函数由式 (6) 给出:

$$L = L_s + \lambda L_{\text{cluster}} \quad (6)$$

式 (6) 为 Softmax loss 与 cluster loss 的加权和, λ 为平衡因子, 用来平衡两种损失函数. 由图 2 所示, 网络由两个输出 Out1 和 Out2, Out1 用来计算 L_{cluster} , Out2 用来计算 L_s .

反向传播过程中, cluster loss 对输入样本 x_i 的偏导数可计算为式 (7), 偏导数 $\frac{\partial L_{\text{cluster}}}{\partial x_i}$ 将反向传播到全局池化层和卷积层.

$$\frac{\partial L_{\text{cluster}}}{\partial x_i} = x_i - c_{y_i} \quad (7)$$

计算第 j 个类的特征中心的差值, 公式为式 (8):

$$\Delta c_j = \frac{\sum_{i=1}^m \delta(y_i = j)(c_j - x_i)}{1 + \sum_{i=1}^m \delta(y_i = j)} + \frac{-\lambda}{\sum_{\substack{i \in N \\ i \neq j}} \sqrt{c_j c_i}} \sum_{\substack{i \in N \\ i \neq j}} \frac{c_i}{2\sqrt{c_j c_i}} \quad (8)$$

其中, 第 1 项是中心损失计算的特征中心的差值, 第 2 项是对 L_{cluster} 的第 2 项求 c_j 的偏导数得出第 j 个类的特征中心的差值. 使用随机梯度下降算法 (SGD) 来对每一批次数据的第 j 类的特征中心进行更新, 学习率是 α , 公式为式 (9). 在算法 1 中, 我们给出了使用 cluster loss 进行识别任务的算法伪代码.

$$c_j^{t+1} = c_j^t - \alpha \Delta c_j^t \quad (9)$$

算法 1. 使用 cluster loss 进行识别任务的算法伪代码

输入: 已经标记的训练数据 $\{x_i, y_i\}$

1. 初始化批次大小为 m , 训练轮次 T , 学习率 μ 和 α , 超参数 λ , 网络权重 ω 和 cluster loss 参数 c_j .
2. for t in range(1, $T+1$)
3. 计算整个网络的损失函数 $L=L_s+\lambda L_{cluster}$
4. 计算反向传播的误差 $\frac{\partial L^t}{\partial x_i^t} = \frac{\partial L_s^t}{\partial x_i^t} + \lambda \frac{\partial L_{cluster}^t}{\partial x_i^t}$
5. 更新 cluster loss 参数 $c_j^{t+1} = c_j^t - \alpha \Delta c_j^t$
6. 更新网络权重 $\omega^{t+1} = \omega^t - \mu \frac{\partial L^t}{\partial \omega^t} = \omega^t - \mu \frac{\partial L_s^t}{\partial \omega^t} - \mu \frac{\partial L_{cluster}^t}{\partial \omega^t}$
7. end for

输出: 网络权重 ω , cluster loss 参数 c_j 和网络损失

2.3 迁移学习

迁移学习, 即让模型具有举一反三的能力. 让网络在任务 A 中学习到的知识应用到任务 B 的学习中, 使得网络不仅有任务 A 的知识也有任务 B 的知识, 任务 A 和任务 B 需有一定的相关性. 文献 [22] 通过大量实验证明了深度学习中迁移学习的可行性. 迁移学习包括 4 种: 样本迁移、特征迁移、模型迁移和关系迁移 [23]. 本文我们使用模型迁移. 如图 3 所示, 首先使用融合图像预训练一个模型, 将预训练模型的 block1, block2, block3 中的所有层的冻结, 不进行权重更新, 然后使用 RGB 图像作为输入, 对预训练模型的其余层进行训练. 之所以冻结 3 个 block 中的网络层, 是因为在卷积神经网络中, 前面几层学习到的都是通用特征, 越靠后的层学习到的特征越具有特殊性 [22].

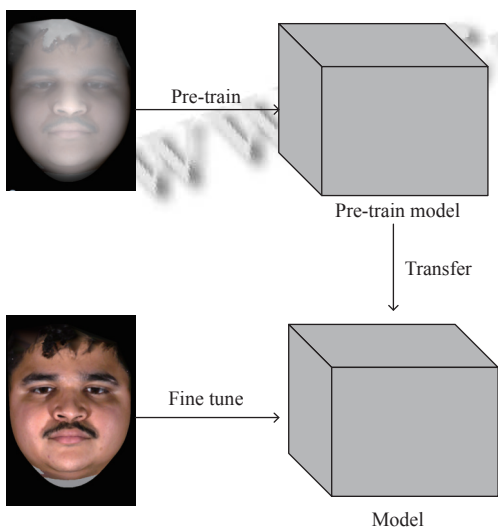


图 3 迁移学习过程

3 实验结果和分析

3.1 数据处理和训练策略

本文实验使用 RGB-D 人脸数据集 Texas3DFRDatabase, 共 118 个人的 1 149 个样本. 使用 TensorFlow 2.0 深度学习框架来搭建网络. 实验室 GPU 型号为 GTX1080Ti. 首先使用第 2.1 节中提到的自动编码器进行 RGB 数据和深度数据的融合, 图 4 分别展示了 RGB 图像、深度图像和融合图像. 对融合图像进行处理, 包括以下几部分.

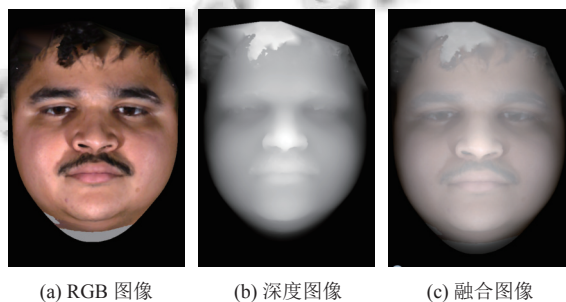


图 4 RGB 图像、深度图像、融合图像

(1) 用 dlib 库自带的 frontal_face_detector 特征提取器进行人脸检测, 并提取出人脸图像.

(2) 将提取出来的人脸图像大小重新设置为 112 像素×112 像素.

(3) 数据增强. 随机对图像进行水平翻转、亮度调整、裁剪、添加噪声, 以防止过拟合问题.

(4) 将增强后的图像进行归一化. 将每一像素的数据类型转换为 float32, 再把每一个像素值除以 255, 使得每一个像素的值在 0-1 之间.

本文实验在使用融合图像预训练模型时, 批量为 64, 共训练 60 轮. 图 2 中 Out1 使用 cluster loss 进行监督训练, Out2 使用 Softmax loss 进行监督训练. λ 设置为 0.07. 前 40 轮使用 Adam 优化器 [24] 对损失函数进行优化, 学习率设置为 0.001, 后 20 轮用 SGD 优化器 [25] 进行优化, 学习率设置为 1×10^{-4} . Adam 的优点是收敛快速, 而正是因为他的优点使得算法有可能错过全局最优解, 因此在前期使用 Adam 加快收敛速度, 后期切换到 SGD, 慢慢寻找最优解. 完成训练后得到一个预训练模型, 然后对预训练模型进行微调. 以 RGB 图像作为输入, 冻结 block1、block2 和 block3 中的所有层, 使用 SGD, 学习率设置为 1×10^{-4} , 对剩余层训练 40 轮, 得到最终模型. 除此之外, 本文分别使用 Softmax loss 和

Softmax loss + center loss 对模型进行预训练,用来和本文提出的损失函数进行对比。

3.2 不同算法性能比较以及 λ 的选择

为了评估有效的评估 cluster loss, 本文在其他因素不变的情况下, 根据不同的损失函数先后训练了 3 个模型: 1) 基于 Softmax loss; 2) 基于 Softmax loss + center loss; 3) Softmax loss + cluster loss. 表 1 展示了 3 种算法训练的结果, 单独使用 Softmax loss 的识别准确率为 92.06%, 加入 center loss 之后, 准确率提升了 2.29 个百分点, 为 94.35%, 使用 Softmax loss + cluster loss 准确率为 97.13%, 提升了 2.78 个百分点. 图 5 为 3 种模型在训练过程中的对比. 单独使用 Softmax loss 收敛最快, 但是准确率相对较低. Softmax loss + center loss 和 Softmax loss + cluster loss 在训练过程中局部有准确率下降的现象, 但是不影响最后结果.

表 1 3 种算法训练结果 (%)

算法	准确率
Softmax loss	92.06
Softmax loss + center loss	94.35
Softmax loss + cluster loss	97.13

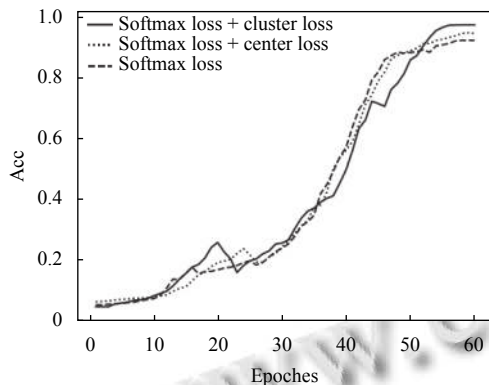


图 5 训练过程对比

超参数 λ 为平衡因子, 用来平衡两种损失函数, 选择合适的 λ 是非常重要的, 能够将两种损失函数达到最佳状态下的平衡, 提高识别准确率. 本文将 λ 控制在 0-0.1 之间, 测试了不同的 λ 对准确率的影响, 实验结果如图 6 所示, λ 值在一定范围内时, 准确率波动不是很大, 相对稳定. 当 $\lambda=0.07$ 左右时, 准确率达到峰值, 因此本文中 λ 选取 0.07.

3.3 测试

本文论述的是三维人脸识别, 前期使用未遮挡、

没有光照变化的人脸图像作为训练集, 因此后期我们需要使用有遮挡、有光照变化的 RGB 人脸图像进行测试来验证本文提出的三维人脸识别算法的有效性. 我们对原始数据集进行了处理, 给人脸加口罩和墨镜模拟了人脸图像在真实条件下的遮挡, 人脸图像调暗模拟光线昏暗的情景, 具体算法本文不详细论述. 处理后的条件如图 7 所示, 分别表示加入口罩、墨镜和光线变暗之后的图像. 处理完之后, 使用处理后的数据分别测试 3 种算法生成的模型, 表 2 展示了测试结果, 限于篇幅原因, S 即为 Softmax, 记录了 3 种算法中不同条件下的识别准确率和平均准确率, 由表可以看出, 3 种算法在不同情景下的准确率相较于使用原始数据训练时都有略微的下降, 这属于正常现象. 本文算法在 3 种情景下都取得了最好的结果, 并且具有很好的鲁棒性, 最后平均准确率高达 96.37%.

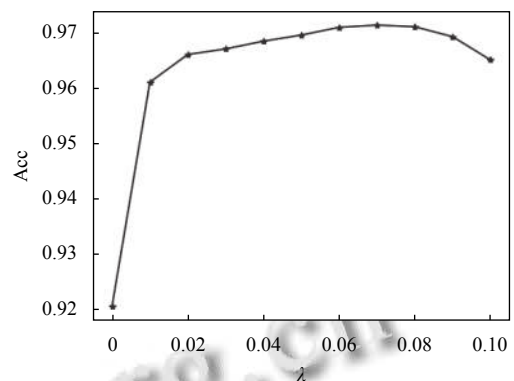


图 6 不同 λ 值对准确率的影响



(a) 戴上口罩 (b) 戴上墨镜 (c) 光线变暗

图 7 加入遮挡、光线处理后的图像

表 2 3 种算法在不同条件下的识别准确率比

算法	口罩	墨镜	光线	平均
S	0.8993	0.9201	0.9009	0.9097
S+center	0.9263	0.9428	0.9390	0.9360
S+cluster	0.9578	0.9707	0.9628	0.9637

3.4 过拟合分析

本实验中, 仍然存在着过拟合现象, 即训练误差与测试误差相差较大, 其原因在于数据量过小. 在算法中

采用了 dropout 和图像增强 (水平翻转、亮度调整, 裁剪、添加噪声) 的方法在一定程度上降低了过拟合现象. 图 8 展示了使用 dropout 和图像增强方法对过拟合的抑制程度. 由图可知 dropout 和图像增强都对过拟合有一定的抑制作用, 两者结合使用效果非常明显, 但是使用图像增强之后, 训练误差上升了一小部分.

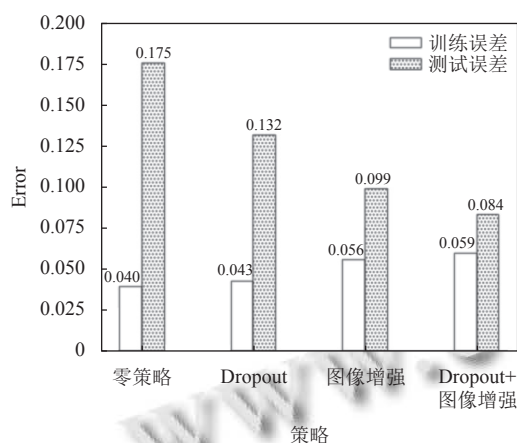


图 8 Dropout 和图像增强方法对过拟合的抑制程度

3.5 多模态有效性实验

为了验证多模态融合特征在三维人脸识别任务中的有效性, 设计了 3 组实验, 分别将融合后的图像、深度图和 RGB 图像作为输入, 进行训练, 3 个模型除输入数据不一样之外, 其余均一致. 训练完成之后, 分别用正常人脸图片、带遮挡人脸图片和暗光线环境下的人脸图片进行测试. 表 3 记录了不同测试环境下的识别正确率.

表 3 多模态有效性实验结果

模态数据	口罩	墨镜	光线	正常
RGB	0.5374	0.5751	0.7361	0.9881
Depth	0.5012	0.5316	0.5289	0.8619
RGB+Depth	0.9578	0.9707	0.9628	0.9853

实验结果表明, 使用多模态融合数据训练出的模型比单独使用 RGB 数据进行训练出的模型, 在正常环境下测试的准确率稍低, 但是在遮挡条件下和光线暗淡条件下, 测试准确率提升非常明显. 由于灰度图质量比较低, 所含信息较少, 因此单独使用灰度图训练出的模型识别精度相对较低. 此实验结果可以证明多模态特征融合的有效性.

4 结论

二维人脸识别受光照、遮挡和姿态的影响较大,

而基于高质量三维扫描数据的人脸识别计算量大, 非常复杂. 针对这一系列问题, 本文基于 RGB 图像和深度图像, 提出了一种多模态融合的三维人脸识别算法. 并且针对人脸数据类间相似性较大的问题, 基于巴氏距离的思想, 提出了一种新的损失函数 cluster loss, 用来惩罚类特征中心的相似性, 使得不同类的特征中心尽可能地分离. 之后使用 Softmax loss、Softmax loss+center loss、Softmax loss+cluster loss 三种算法分别训练了模型进行比较, 用来验证我们提出的损失函数的优越性. 我们使用处理过的 RGB 图像 (模拟真实情况下口罩和墨镜对人脸的遮挡和光线较暗的情景) 对预训练的模型进行微调, 使得模型仅使用 RGB 图像作为输入即可, 并且不受光照和遮挡的影响. 测试表明, 我们的算法受光照和遮挡的影响非常小, 并且我们提出的损失函数在测试中较其他两种取得了最优的效果, 识别准确率达到了 96.37%. 但是本文提出的方法具有一定的局限性, cluster loss 需要和 Softmax loss 联合使用, 超参数 λ 的根据经验设置, 没有具体的公式来量化. 我们仅考虑了闭集测试, 没有考虑开集测试. 在下一步的工作中, 我们会继续优化算法, 并且使用开集测试.

参考文献

- 1 Savran A, Alyüz N, Dibeklioglu H, *et al.* Bosphorus database for 3D face analysis. The 1st European Workshop on Biometrics and Identity Management. Roskilde: Springer, 2008. 47–56.
- 2 Yin LJ, Wei XZ, Sun Y, *et al.* A 3D facial expression database for facial behavior research. 7th International Conference on Automatic Face & Gesture Recognition. Southampton: IEEE, 2006. 211–216.
- 3 Mu GD, Huang D, Hu GS, *et al.* Led3D: A lightweight and efficient deep approach to recognizing low-quality 3D faces. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019. 5766–5775.
- 4 Zhang JJ, Huang D, Wang YH, *et al.* Lock3DFace: A large-scale database of low-cost Kinect 3D faces. 2016 International Conference on Biometrics (ICB). Halmstad: IEEE, 2016. 1–8.
- 5 Min R, Kose N, Dugelay JL. KinectFaceDB: A kinect database for face recognition. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2014, 44(11): 1534–1548. [doi: 10.1109/TSMC.2014.2331215]
- 6 Gupta S, Castleman KR, Markey MK, *et al.* Texas 3D face

- recognition database. 2010 IEEE Southwest Symposium on Image Analysis & Interpretation. Austin: IEEE, 2010. 97–100.
- 7 Gilani SZ, Mian A. Learning from millions of 3D scans for large-scale 3D face recognition. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 1896–1905.
- 8 Cai Y, Lei YJ, Yang ML, *et al.* A fast and robust 3D face recognition approach based on deeply learned face representation. *Neurocomputing*, 2019, 363: 375–397. [doi: [10.1016/j.neucom.2019.07.047](https://doi.org/10.1016/j.neucom.2019.07.047)]
- 9 Kim D, Hernandez M, Choi J, *et al.* Deep 3D face identification. 2017 IEEE International Joint Conference on Biometrics (IJCB). Denver: IEEE, 2017. 133–142.
- 10 皮家甜, 杨杰之, 杨琳希, 等. 基于多模态特征融合的轻量级人脸活体检测方法. *计算机应用*, 2020, 40(12): 3658–3665.
- 11 Kusuma GP, Chua CS. PCA-based image recombination for multimodal 2D + 3D face recognition. *Image and Vision Computing*, 2011, 29(5): 306–316. [doi: [10.1016/j.imavis.2010.12.003](https://doi.org/10.1016/j.imavis.2010.12.003)]
- 12 Jiang L, Zhang JY, Deng BL. Robust RGB-D face recognition using attribute-aware loss. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(10): 2552–2566. [doi: [10.1109/TPAMI.2019.2919284](https://doi.org/10.1109/TPAMI.2019.2919284)]
- 13 Lee YC, Chen JC, Tseng CW, *et al.* Accurate and robust face recognition from RGB-D images with a deep learning approach. *Proceedings of the British Machine Vision Conference 2016*. York: BMVA Press, 2016. 1–14.
- 14 Li BYL, Xue ML, Mian AS, *et al.* Robust RGB-D face recognition using Kinect sensor. *Neurocomputing*, 2016, 214: 93–108. [doi: [10.1016/j.neucom.2016.06.012](https://doi.org/10.1016/j.neucom.2016.06.012)]
- 15 Chang KI, Bowyer KW, Flynn PJ. An evaluation of multimodal 2D+3D face biometrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, 27(4): 619–624. [doi: [10.1109/TPAMI.2005.70](https://doi.org/10.1109/TPAMI.2005.70).]
- 16 Cui JY, Han H, Shan SG, *et al.* RGB-D face recognition: A comparative study of representative fusion schemes. 13th Chinese Conference on Biometric Recognition. Urumqi: Springer, 2018. 358–366.
- 17 Wen YD, Zhang KP, Li ZF, *et al.* A discriminative feature learning approach for deep face recognition. 14th European Conference on Computer Vision. Amsterdam: Springer, 2016. 499–515.
- 18 Deng JK, Guo J, Xue NN, *et al.* ArcFace: Additive angular margin loss for deep face recognition. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 4685–4694.
- 19 Cai J, Meng ZB, Khan AS, *et al.* Island loss for learning discriminative features in facial expression recognition. 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition. Xi'an: IEEE, 2018. 302–309.
- 20 Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv:1511.06434, 2015.
- 21 Shi WZ, Caballero J, Theis L, *et al.* Is the deconvolution layer the same as a convolutional layer? arXiv: 1609.07009, 2016.
- 22 Yosinski J, Clune J, Bengio Y, *et al.* How transferable are features in deep neural networks? *Proceedings of the 27th International Conference on Neural Information Processing Systems*. Montreal: MIT Press, 2014. 3320–3328.
- 23 Pan SJ, Yang Q. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 2010, 22(10): 1345–1359. [doi: [10.1109/TKDE.2009.191](https://doi.org/10.1109/TKDE.2009.191)]
- 24 Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv: 1412.6980, 2014.
- 25 Bottou L. Stochastic gradient descent tricks. In: Montavon G, Orr GB, Müller R, eds. *Neural Networks: Tricks of the Trade*. 2nd ed. Berlin: Springer, 2012. 421–436.

(校对责编: 孙君艳)